UNITED NATIONS
ECONOMIC AND SOCIAL COUNCIL

E

Distr.
GENERAL

E/ECA/PSD.4/57
12 December 1985

Original: ENGLISH

ECONOMIC COMMISSION FOR AFRICA

Fourth Session of the Joint Conference of African Planners, Statisticians and Demographers

Addis Ababa, Ethiopia, 3-12 March 1986

LC/NAIROBI.

ADJUSTMENT OF ERRORS IN THE REPORTED AGE-SEX DATA FROM AFRICAN CENSUSES

C O N T E N T S

		Page
I.	INTRODUCTION	1
II.	COLLECTION OF DATA ON AGE AND SEX	1-3
III.	QUALITY OF AGE-SEX DATA	3-4
IV.	ADJUSTMENT OF AGE-SEX DATA	4-9
V.	CONCLUDING REMARKS	9

I. INTRODUCTION

1. Population is central in the planning process. Not only do we need the size of a population, but also its age-sex structure because errors in these parameters could lead to faulty planning. For example, the health needs are very much different for a young than for an old population. In estimating the needs of a population in regard to education, health, labour force, housing and food, the planner and policy maker will have to depend on pertinent population statistics classified by age and sex. Besides, the demographic and socio-economic situation of a country is determined by its age-sex structure. The effect of age-sex composition on the fertility of a population is also obvious; only women in the childbearing ages of 10 and 54 can contribute to fertility. Many demographic socio-economic attributes are sex specific. Data classified by sex not only have analytical importance, they could also be used as evaluative tools and sex classification is essential in certain situations because of biological differences. Migration also is age-sex selective. Since participation rates in education and economic activity may also vary between age and sex groups, the age-sex structure of a population is important in the study of these important aspects.

2. Although the knowledge of the size, composition, dynamics and distribution of a population is essential for planning, changes in these parameters over time and space are also pertinent. In this regard, if the age-sex data at a base period is subjected to errors and deficiencies, then any projection will not only carry forward these deficiencies, but would also be affected by the effects of such errors which will show up in estimation of births, deaths and migration.

3. The majority of the ECA member States lack direct information on levels and trends of fertility, mortality and migration; data on age-sex composition of the population are mostly utilized to estimate these parameters. With some of the indirect methods of fertility and mortality estimation, the effects of age-sex errors on the resultant parameters are quite large and can vitiate conclusions.

4. Thus the need to have accurate age-sex information of a population need not be over-emphasized. There are two ways of improving the age-sex data. The first one is direct and involves measures to ensure better quality information from the field. Use of aids like historical calendars and documents are some of the techniques available. The other way is to adjust the collected data by various techniques. This paper will address itself mostly to the various adjustment techniques of errors in reported age-sex data; but mention will be made of how and why errors occur in reported age-sex data so that the data collection agency in the member States could plan strategies to counteract the problems and ensure improved age-sex data.

II. COLLECTION OF DATA ON AGE AND SEX

5. The measurement of age has several problems. Information on age can be collected in completed years or in ages nearest to next birthday. The question can be asked; what is a year? To one who is familiar with the western concept of age, it connotes a solar year. But to a Muslim, it means a lunar year; and, to a Chinese it may refer to the traditional calendar. To yet others, it may be quite different from all these.

Similar problems arise regarding the meaning of a birthday. Very few people in the member States celebrate birthdays as relatively little importance is put on the event. Even though there are tremendous variations in age reckoning among societies, it is possible to convert (approximately) age in one system tothat in another. Thus the real problem in age estimation is not the different systems adopted, it is ignorance or indifference as to when a person is born. Deliberate misstatements of age due to one reason or the other, is yet another dimension and this could create problems.

6. When ages are not known to respondents, they are estimated either by them or by others (including the enumerator). The first approximation is to put persons into one of the appropriate decennial groups; thus a large number of people are reported as aged 10, 20, 30, ... years. When efforts are made to estimate age more closely then the mid-point of the range is used and this culminates in persons being reported as aged 15, 25, 35, ... years. Thus digits 0 and 5 become preferred digits in that order. Furthermore, approximation within five year ranges brings in reported ages ending in digits 2 and 8 with minor peaks at 4 and 6. This is also brought about by avoidance of odd digits and preference for even digits. This kind of digit preference or age heaping has been observed in most member States. However, the tempo of age heaping varies over the age range and becomes more and more at the advanced ages so much so that very few people are reported at ages other than 60, 70, 80, etc. In Sudan, for instance, at the 1973 census, around 50 per cent of persons were reported with ages ending in 0 and 5 as against 60 per cent in the age group 60-69.

7. In spite of all efforts to estimate ages using historical or other calendar of events, physical appearance, physiological changes, social customs like circumcision and age grading, documents (e.g. civil and resident registration), it is still possible that a large number of cases remain where age estimation is impossible. As long as the resulting age not stated or unknown group is not large (less than 2 per cent) there is not much of a problem; in most cases prorata allocation may suffice. If other information are available to allocate these into specific age groups, then this could be used. Cold and hot deck methods of imputation are well known. But imputation may create problems in data analysis, interpretation and utilisation and hence should be avoided as much as possible.

8. One of the commonest observations in the reported ages of the member States is the relative shortage of infants and children in enumeration. Deliberate omission, misunderstanding of the scope of enumeration and age estimation error could be possible reasons. Some duplication could result from de jure counting in extended family systems.

9. Even though the definition and classification of sex are easy and sex is generally ascertainable, still field data have their due share of problems arising from misclassification of data by sex. There are several reasons why one sex of the other is either misreported, wrongly reported or not fully reported or even over-reported. At certain ages, the reporting errors could be sex spedific. For instance, omission of infants and children could different from one sex to another. Whereas omission of male children have been noted in most member States due perhaps to fear of "evil eye", the opposite could happen if female children are not considered important to be reported. Ignorance could play an important part in these. Again, reporting of age of males and females differs when they reach puberty and post puberty ages.

Young unmarried girls usually are reported as younger and those married and/or have children are reported as older, but with a noted tendency to be within the reproductive age. For males the tendency is to exaggerate age of adults. Deficit of males in the reproductive ages could also be the result of their mobility, migration or work status. Whereas some omission of girls of marriageable age have been reported among Muslim or Arab-influenced populations, omission of adult males in the past were mostly due to imposition of hut tax and military conscription.

10. In addition to reporting errors, there are instances where recording errors have been detected. A common observation has been non-recording of the sex of a person probably because to the enumerator, the name would automatically suggest the sex. There is also the problem of legibility in writing. For instance, numerals like 1 and 2, unless written with care, may get distorted at the further processing stage. When ages of infants are to be reported in months, this should be clearly marked, as otherwise a 9-month old child may be recorded as 9 years old, and vice versa. No blanks should be left in the questionnaire. In this connection there is need to ensure that after the data has been brought from the field, further errors and biases are not introduced at the processing, analysis and publication stages; all coding should be verified to ensure error free data.

III. QUALITY OF AGE-SEX DATA

11. In the 1980 round of censuses, available information indicates that many of the member States endeavoured to collect age data in single years of age. This represents some improvement over the earlier practice of using such broad age groupings as infants, young children, adults and old persons. The percentage reported as 'age not stated' is also low. But these do not imply that automatically the quality of the data has improved or that people have acquired better information about their ages. Of course the scope of the data collected by these States has improved but the data quality is still relatively poor.

12. An accurate enumeration of the population classified by age involves many difficulties. In general inadequacies in age statistics arise from two basic sources - failure to report ages and deliberate misstatements of ages. The latter may arise from ignorance of the correct age; a tendency to understate at some ages while exaggerating at others - particularly at the more advanced ages; either a conscious or subconscious preference for certain socially significant ages and a corresponding avoidance of other ages; a general tendancy to over-select ages ending in certain digits and, deliberate falsification of age reports for a variety of social, economic, political or purely individual motives. Thus in the reporting of age from censuses there arise five major forms of content errors and biases. These are the under-reporting of the number of the children aged under one; a tendency to give exact age of some legal significance (e.g. voting at elections, marriage, etc.); distinct overstatement of age among those at very advanced ages; the reporting of some individuals as being of an unknown age; age heaping. 1/

1/ Myers, R.J., "Accuracy of age reporting in the 1950 United States Census", (JASA, 1954), vol. 49, pp. 826-31.

13. There are several methods of appraising the quality of age-sex data. It is pertinent to observe at this point that the types of adjustment techniques used should be related to the data quality. A detailed appraisal of these various techniques has been undertaken in an earlier work. 2/ This paper will focus on two of the indices used to detect errors in age heaping. The two indices are those by Myers for evaluation of single years age-sex data and that by the United Nations (i.e. the Joint Score Index) for evaluation of 5 years age-sex data. Based on the joint score index, a quality rating is inferred. The reported age-sex data for a member State is presumed accurate if the joint score index is between zero and twenty; inaccurate if the score is between 20 and 40; and, highly inaccurate if the score is above 40. These are denoted by symbols A, B and C respectively.

14. Table 1 presents data on both indices for 39 member States with pertinent age-sex data from the most recent census. Only one member State (Algeria) fell into the A grade category while 12 had scores of at least 60 implying very poor data. However, from the joint score index computed at two points in time for four of the member States (Table 2), this quality is improving.

IV. ADJUSTMENT OF AGE-SEX DATA

15. In deciding on the adjustment procedure to be adopted in any particular case, some basic principles should be kept in mind. Any adjustment of the data should take care of the peculiarities in the data so that these are not submerged under ad hoc adjustment techniques. Moreover, the adjusted data should not only be smooth, it should also be consistent with the demographic, socio-economic realities of the country both in the past and at the point of time in question. Moreover, adjustment should also ensure smoothness in three directions - vertical (age ratios), horizontal (sex ratios) and diagonal (survival ratios). This means that the consecutive age-sex groups and survival ratios should all be within the realm of possibilities. It is equally important to keep the adjustment to the minimum and in particular, to ensure that the procedures adopted are consistent with the types, varieties and magnitudes of the observed errors and deficiencies.

(i) Mathematical Methods

16. Early in the course of developing procedures for adjusting errors in reported age-sex data, the United Nations proposed differential methods for adjusting the errors in the age groups 0-4, 5-9, 10-74 and 75+. 3/ As advances were made in

---

2/ ECA, "Analytical techniques for content error evaluation", in Statistical Bulletin for Africa No. 14, E/CN.14/SIB/14, pp. 79-117.

3/ For details of the methods see Method for Population Projections by Sex and Age, Manual 3 (United Nations, 1956), pp. 11-14.

demographic techniques, mathematical formulae were developed for handling the problem. These techniques included the Carrier-Farrag ratio, Newton's formula and Osculatory graduation method; an outline of their description4/ is presented in Annex 1.

17. The application of these mathematical formulae in varying circumstances has led to the general conclusion that none of the methods is fool proof. In particular, the United Nations approach can be labourious in a situation where the analyst has to deal with data for many member States. Besides, the 5-point moving average method used in smoothing errors in ages 10-74 by the United Nations method as well as the mathematical formulae by Newton, Carrier-Farrag and Osculatory graduation simply redistribute the population in a smooth fashion without taking account of the demographic nature of the data. Therefore, if the observed distortions in the age distribution at a particular point in time reflect unique historical demographic trends, by simply smoothing the observed age distribution, these mathematical approaches do not take the nuances of the data into consideration.

18. Besides, the types and magnitude of the errors and biases may not be the same or even similar from one age group to the next or between two sex groups or even over a time period. Thus one mathematical formula might compromise some of the nuances in the data and at the most, ensure vertical consistency (age ratios). Horizontal smoothness (sex ratios) and diagonal consistency (survival ratios) might not be embedded in these smoothing procedures.

19. Tables 3 to 7 present the adjusted age-sex distributions using the Carrier-Farrag, Newton's quadratic and Sprague osculatory methods for Morocco (1971), Ghana (1970), United Republic of Cameroon (1976), Kenya (1969) and Lesotho (1976). As it is obvious from these data, all these methods at the most can tackle only the population aged 10-69 but not the very young (0-4) and old persons (70+), where the errors and biases are more pronounced and especially at the young ages where the need for adjustment is more pressing than at other ages.

20. As noted earlier, the United Nations in their very early attempts at adjustment of data, recommended special methods to adjust the very young and older ages. For the children aged 0-4 years, the United Nations method suggested is based on the "reverse survival ratio method" by which the child population is projected backwards for five years to derive the implied birth rate and verify whether the two are consistent (i.e. the projected population 0-4 and the associated level of fertility).

21. In one version of the method, the entire population is reverse survived using an appropriate life table for the past two five-year periods. The estimated births are converted into sex-age adjusted birth rates and compared with each other and with expected values. Adjustment is then carried out on ages 0-4 and 5-9 to reflect the fertility levels. This method is less affected by errors in the selection of life tables and is preferred in cases where natural increase is the predominant factor of population growth. But when migration might have played a part in population change, then only the first two five-year age groups are projected backwards and the entire population is reversed for five or ten years by applying appropriate growth rates.

22. For adjusting the population at older ages, the practice is to compare the total proportion at ages 70 years and above with a model stationary population; the model percentages at five-year groups are read off and used to adjust the observed total. If age exaggeration occurs at earlier ages, then it is advisable to consider ages 60-64, 65 and above; if some idea of the fertility and mortality conditions is known, then a model stable population is preferable to a stationary population.

4/ See Carrier, D.H. and Farrag, A.M., "The reduction of errors in census population for statistically underdeveloped countries", Population Studies (March, 1959), pp.240-285.

(ii) Demographic methods

23. The need to take account of the nuances in the data led to the application of analytical techniques such as demographic models whereby the age-sex distribution of the population in question is assumed to fit an appropriate population model selected on the basis of associated demographic parameters and other relevant information.5/ There are several models to choose from including the United Nations model, the Coale-Demeny West/North/South/East models, the Brass logit and Lederman systems. Recently, the United Nations published a family of model life tables and stable populations which it designates as - Latin American, Chilean, South Asian, Far Eastern and general patterns. African data was rarely used in the derivation of most of these. From experience it has been suggested that sub-Saharan Africa is near the North model of Coale-Demeny and North African countries are near the South model.6/

24. The selection of a model may not be the best solution to a particular situation and adjustment of the data might be based on some modification of the above procedure. For instance, the model may be accepted for certain portions of the age-sex distribution and other portions may be adjusted by ad hoc methods which take special care of the peculiarities of the data on hand as is the case with the United Nations method described earlier. Since demographic parameters might undergo some change over time, the use of quasi-stable population models might be recommended in such instances. Equally, because the stable or quasi-stable population might not reflect some of the nuances in the data of a member State at a particular point in time, a combination of model population and mathematical methods has been suggested by Brass.

25. In his logit difference method, to obtain an adjusted age-sex (in five-year groups) population distribution from that reported in a census, he proposes the fitting of a straight line to the logit of differences between the reported age distribution and that of a reference standard. The intercept (a) and the gradient (b) of the fitted lines can be derived by using the relation.

$$\text{logit } P_o(x) = a + b \text{ logit } P_s(x) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

where $P_o(x)$ = reported percentage values at age x,

$P_s(x)$ = reference standard percentage at age x.

5/ United Nations, Methods of estimating basic demographic measures from incomplete data, (United Nations, New York, 1967), pp. 17-27.

6/ Clairin, R., "The assessment of infant and child mortality rates from different sources", in the Population of Tropical Africa (Eds. Caldwell, J.C. and Okonjo, C.), Longmans (1968). See also Ekanem, I.I. and Som, R.K., "Problems of choosing model life tables for African countries"; GENUS, vol. XL-niv-2(1984), pp. 57-70.

In order to solve for the values of a and b in (1), use can be made of the following normal equations:

$$\sum_1 {}_1P_o(x) = an + b \sum_1 {}_1P_s(x) \quad \ldots\ldots(2)$$

and $$\sum_2 {}_2P_o(x) = an + b \sum_2 {}_2P_s(x) \quad \ldots\ldots(3)$$

where n = number of observations in each group,

$\sum_1$ and $\sum_2$ = summation over first and second groups respectively.

26. Because of the observed tendency to exaggerate ages at older age groups, the method excludes proportions above age 60. For the purpose of calculation, the observed and reference standard age distributions at 5 year intervals from 5 to 60 years inclusive are separated into two equal groups arranged in order of age. The first group consists of proportions in ages 5, 10, 15, 20, 25 and 30 while the second consists of 35, 40, 45, 50, 55 and 60. The procedure is sometimes repeated for ages 5 to 50 (Group 1: 5, 10, 15, 20 and 25; Group 2: 30, 35, 40, 45 and 50) and ages 5 to 40 (Group 1: 5, 10, 15 and 20; Group 2: 25, 30, 35 and 40) to choose the best fit in any given situation.

27. However, there is need for caution in the use of the age distribution derived on the basis of the Brass logit difference approach. This is because, in retrospect, the intercept (a) and slope (b) derived on the basis of the reported and reference standard age distributions are used to adjust the reported age distribution. In effect, as in the case of the quasi-stable model approach, the same problem of subjectivity in selecting the stable population parameters is still there with the logit approach. The main difference between the two lies in the fact that whereas with the quasi-stable approach the reference age distribution is borrowed without any adjustment and hence may not reflect the peculiarities in the data; the logit approach makes room for such an adjustment. To this extent, therefore, it is an improvement over the quasi-stable approach.

28. The advantage claimed in using the Brass logit method is that some of the peculiarities of the observed distribution are reflected in the fitted distribution unlike the stable population model method. The question here is whether this is really an advantage, when some of the peculiarities of the data may be due to the very errors which the analyst seeks to remove or at least minimise in the adjusted age-sex data. This contention is illustrated with examples of adjusted age-sex data from a few member States.

29. In the evaluation and analysis of the 1976 census age-sex data of Swaziland, the Brass logit method was applied to graduate the data and the results in Table 8 are given for a portion of the age distribution only to illustrate the point. From the data, one easily notes from the adjusted and projected proportions (0-4), that the adjusted percentage at age 0-4 for 1981 is not consistent with the estimated fertility parameter. The basic shape of the graduated distribution was dictated by that of the reported distribution which comprised a rather smaller 1981 proportion of children (17.6 per cent) than would be implied (19.0 per cent) by the estimated fertility and mortality rates. Thus the projected figures show a slight irregularity at the point where the new birth cohorts are fed in.

30. Another observation based on the projection of the Swaziland population is related to the sex ratio of the population. Whereas at base date (1976) it was taken as 95.0, it increased to 95.1 in 1981 and 95.2 in 1986 contrary to what one might have expected from the improvement in mortality which already had a great advantage for the females. The actual problem came from the child population whose sex ratio played a great role in determining the sex ratio of the population. As time passes, the sex ratio of a population is much influenced by the sex ratio of the child population and their survivors. If the sex ratio at birth is taken as high or low, the effect would be reflected on the overall sex ratio. In this particular case perhaps the assumed sex ratio at birth of 103 for the projections seems too high and may be only 102 or even 101.

31. A similar experience is noted in the case of Botswana where the logit method produced an adjusted population distribution which was at variance with the implied fertility-mortality estimates. Since the data showed not much age-sex reporting errors, the UN method of adjusting the population 0-4, the adult and old age groups resulted in a distribution which tallied with the fertility-mortality conditions. Table 9 gives the results obtained.

32. After adjusting the age-sex data, it is very difficult to assert that it is the best under the circumstances because one does not know what the true age-sex distribution of the population is without the inherent errors, biases and deficiencies. A rule of thumb is that it is better to apply a battery of adjustment techniques and search for the one which deviates minimally from the observed or estimated fertility-mortality situations. Several forward and backward projections need be carried out using the age-sex distribution and the fertility-mortality estimates and the performance and consistency of the adjusted data be studied.

33. To illustrate the use of the Brass logit technique for smoothing reported age-sex distributions needed for population projections, the results obtained in applying the technique to the age-sex data from the 1976 census of Egypt are reproduced in Table 10. The derived total population by sex for 1980 was distributed by age using the smoothed 1976 distribution. This was projected backwards and forwards on the basis of assumptions regarding the growth components. The resulting sex ratios for 1950 clearly indicate the fact that there were problems with the smoothed age distribution adopted for the base year 1980.

34. When age-sex data are adjusted by some of the techniques mentioned earlier, it may so happen that the sum of the adjusted population in the various age groups may differ from the enumerated figure. The question then is: should the figure be prorated to the enumerated population or should one keep the adjusted figures? Even though the adjustments may be really necessary in order to remove the errors, deficiencies and biases in the age-sex data, since the census reported population is generally better known and accepted by the country and others and moreover, different evaluations and adjustments may result in varying adjusted population figures, it is advisable to keep the total population by sex same as the census enumerated figure and prorata adjust the age distribution accordingly. However, if a different base year or period is utilised, especially while carrying out a projection exercise, it is preferable to utilise the adjusted populations by age and sex as also the total population so obtained and carry it forward or backward for the required length of time with estimated growth rates in order to obtain the needed base age-sex distribution.

35. Besides the foregoing indirect adjustment techniques, errors in reported age-sex data can be adjusted directly, Because age-sex data from post-enumeration surveys (PES) are often based on more carefully planned and collected information, they may be used to adjust the age-sex distribution from the main census. Table 11 presents the effects of adjusting the male age data from the 1974 census of Liberia. It can be seen that not only did the age distribution become closer to a stable model, the estimated birth rates also are closer. The sum of squares of deviations from unity of the ratios of stable to the observed is reduced. Thus if PES or other extraneous information of a better quality were available, it seems useful to use these to adjust observed age-sex distribution of population and any other types of deficiencies in the data.

V. CONCLUDING REMARKS

36. Adjustment techniques have not yet reached a stage where routine techniques can be applied without other tests. Fortunately, most of the adjustment techniques and forward and backward projections are computerized and often these techniques can even be carried out with a programmable electronic calculator. The demographer analyst should devote enough time and attention to study the results of these and only then select the final adjusted distribution.

37. Adjustment of data is still an art and experience and knowledge of the population is very essential in carrying out a realistic adjustment. Sometimes one is caught in the vicious circle of quality of data biasing estimates of fertility and mortality and biased estimates of fertility and mortality misleading the analyst regarding the adjustments needed. Since prevention is better than cure, it is strongly recommended that more efforts be put in improving the collection of the data from the field. Education, use of aids in estimating ages, better enumerators, stricter supervision, quality controls at all stages of data processing, etc. would yield dividends in terms of improved quality of the data.

38. No amount of adjustment can improve a given data and any pains taken to assure better quality data from the field and guarding against further deterioration in the processing of the collected data would be worth the while. It is gratifying to note that the quality of data has certainly improved in the member States where a comparison could be carried out. As many of these States had their first experience during the 1980 round only, the opportunity to improve the basic data collected and processed depends on the statistical office and systems in the member States concerned.

Table 1: Distribution of Age and Sex ratio scores, Joint score index, Myer index and data quality rating for selected ECA member States

Country	Census Year	Age ratio Score M	Age ratio Score F	Sex ratio Score	Joint Score Index a/	Quality Rating	Myers Index M	Myers Index F
NORTHERN								
Algeria	1977	5.3	5.5	3.2	20.2	A		
Egypt	1976	10.2	14.2	8.1	48.7	C		
Libyan A. Jamahiriya	1973	5.6	11.0	6.1	34.9	B	9.8	25.8
Morocco	1971	14.2	25.2	16.2	88.0	C		
Sudan	1973	18.3	20.6	8.7	65.2	C	48.2	58.0
Tunisia	1975	6.9	5.2	5.3	28.0	B		
WESTERN								
Burkina Faso	1975	10.1	14.1	10.4	54.5	C		
Cape Verdé	1980	15.5	13.3	6.4	48.0	C		
Gambia	1983	15.1	24.9	13.4	80.2	C	42.9	50.7
Ghana	1970	8.3	11.6	7.6	42.7	C	26.8	31.2
Ivory Coast	1975	5.8	9.2	10.6	46.8	C		
Liberia	1974	11.2	14.6	11.8	61.2	C	27.5	29.4
Mauritania	1977	9.6	25.1	14.3	77.6	C		
Mali	1976	9.6	19.5	12.2	65.6	C		
Niger	1977	26.6	47.0	26.7	153.7	C		
Nigeria	1963	29.0	36.4	14.1	107.7	C		
Senegal	1976	4.8	9.7	14.2	56.9	C		
Sierra Leone	1974	13.2	16.1	8.5	54.6	C	44.4	47.2
Togo	1970	18.7	19.0	11.6	72.5	C		
CENTRAL								
C.A. Republic	1975	18.9	26.2	9.7	74.2	C		
Congo	1974	4.9	4.1	5.2	24.7	B		
Gabon	1960	11.4	13.2	10.0	54.5	C		
Sao Tomé & Principé	1981	7.3	5.4	5.5	29.2	B		
U.R. of Cameroon	1976	5.6	6.8	5.6	29.2	B		
EASTERN								
Burundi	1979	7.2	10.6	5.2	33.4	B		
Djibouti	1983	15.9	20.5	9.9	65.9	C		
Kenya	1979	3.4	4.0	4.2	20.8	B	12.4	15.4
Madagascar	1975	4.0	5.8	5.0	24.8	B		
Malawi	1977	14.4	9.2	8.3	48.4	C	12.9	16.6
Mauritius	1983	11.5	8.4	4.4	33.1	B	2.1	2.3
Mozambique	1980	10.9	12.6	8.0	47.4	C		
Rwanda	1978	6.6	6.5	3.7	24.1	B		
Seychelles	1980	5.8	6.1	5.9	29.5	B		
U.R. Tanzania	1978	10.1	8.8	9.1	46.1	C	25.2	32.1
Uganda	1969	8.5	12.4	8.9	47.7	C	24.6	33.3
Zambia	1980	14.8	8.3	14.7	67.2	C	14.1	15.3
SOUTHERN								
Botswana	1981	3.0	5.3	5.3	24.1	B	7.1	6.8
Lesotho	1976	9.2	8.7	5.3	33.8	B	10.4	9.6
Swaziland	1976	8.9	6.2	5.6	31.9	B	14.1	14.7

Source: Various country distributions of single and five year of age by sex data for census year.

a/ Based on ages 0-69 in each case by sex.

Table 2: Age ratio, Sex ratio and Joint Scores for selected countries from their recent two censuses

Country	Year	Age ratio score M	Age ratio score F	Sex ratio Score	Joint Score	Myers Index M	Myers Index F
Ghana	1960	12.4	14.4	8.8	53.2	29.6	31.2
	1970	8.3	11.6	7.6	42.7	26.8	31.2
Kenya	1969	5.1	7.4	6.7	32.6	21.2	22.2
	1979	3.6	4.0	4.2	20.2	12.4	15.4
Lesotho	1966	10.0	9.7	6.5	39.2	11.8	12.8
	1976	9.2	8.7	5.3	33.8	10.4	9.6
Libya	1964	13.4	21.1	14.4	77.7	46.5	70.3
	1973	5.6	11.0	6.1	34.9	9.8	25.8

Table 3: Reported and smoothed age-sex distributions and indices of dissimilarity for Morocco, 1971

Age Group	Reported		CF Smooth		QD Smooth		OS Smooth	
	M	F	M	F	M	F	M	F
0-4	16.5	16.3	16.4	16.1	16.4	16.1	16.7	17.2
5-9	16.4	16.0	16.4	15.9	16.4	15.9	16.2	14.9
10-14	14.5	13.0	13.7	12.3	13.5	12.3	13.8	12.4
15-19	9.8	9.3	10.6	10.0	10.9	10.0	10.5	9.9
20-24	6.6	7.2	6.6	7.5	6.8	7.6	6.9	7.5
25-29	5.3	6.6	5.3	6.4	5.1	6.3	5.0	6.3
30-34	5.1	6.5	5.4	6.4	5.3	6.4	5.2	6.3
35-39	5.1	5.5	4.9	5.6	4.9	5.7	5.0	5.7
40-44	4.6	5.1	4.4	4.6	4.3	4.5	4.4	4.6
45-49	3.5	3.0	3.7	3.6	3.8	3.6	3.7	3.6
50-54	3.5	3.2	3.0	2.5	3.0	2.6	3.0	2.6
55-59	2.0	1.3	2.5	2.1	2.5	2.0	2.4	2.0
60-64	2.3	2.6	2.0	2.0	2.0	2.0	2.0	1.9
65-69	1.3	1.1	1.6	1.7	1.6	1.7	1.6	1.8
70-74	1.6	1.7	1.6	1.7	1.6	1.7	1.3	1.5
75-79	0.6	0.5	0.6	0.5	0.6	0.5	1.0	0.7
80+	1.3	1.1	1.3	1.1	1.3	1.1	1.3	1.1
Index of dissimilarity with reported	-	-	1.95	3.10	2.60	3.20	2.55	4.15

CF = Carrier-Farrag ratio method

QD = Newton's quadratic method

OS = Osculatory graduation method

Table 4: Reported and smoothed age-sex distribution and indices of dissimilarity for Ghana, 1970

Age Group	Reported		CT Smooth		QD Smooth		OS Smooth	
	M	F	M	F	M	F	M	F
0-4	18.1	18.3	18.2	18.2	18.4	18.3	20.0	20.6
5-9	17.5	16.7	17.1	16.7	17.1	16.7	15.5	14.5
10-14	12.0	11.3	12.0	11.0	12.1	11.2	12.1	10.9
15-19	9.4	8.8	9.5	9.1	9.4	8.9	9.4	9.2
20-24	7.2	8.7	7.6	8.9	7.6	8.8	7.5	8.7
25-29	6.8	7.9	6.5	7.8	6.4	7.8	6.5	7.9
30-34	6.2	6.9	6.2	6.6	6.1	6.6	6.1	6.6
35-39	5.2	5.0	5.3	5.3	5.3	5.4	5.3	5.3
40-44	4.1	4.1	4.2	4.0	4.2	4.0	4.2	4.0
45-49	3.4	3.0	3.3	3.1	3.3	3.0	3.3	3.0
50-54	2.8	2.6	2.6	2.3	2.6	2.3	2.6	2.3
55-59	1.8	1.5	2.0	1.8	2.0	1.8	2.0	1.8
60-64	1.8	1.6	1.7	1.5	1.6	1.5	1.6	1.5
65-69	1.1	1.1	1.2	1.2	1.3	1.2	1.3	1.2
70-74	1.0	0.9	1.0	0.9	1.0	0.9	0.9	0.9
75-79	0.5	0.5	0.5	0.5	0.5	0.5	0.6	0.5
80+	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1
Index of dissimilarity with reported	-	-	1.20	1.30	1.30	1.05	2.95	3.40

Table 5 : Reported and smoothed age-sex distribution and indices of dissimilarity for United Republic of Cameroon, 1976

Age group	Reported		CF Smooth		QD Smooth		OS Smooth	
	M	F	M	F	M	F	M	F
0-4	17.5	16.6	17.4	16.6	17.4	16.6	18.2	17.6
5-9	15.6	14.6	15.8	14.6	15.6	14.6	15.0	13.6
10-14	12.1	10.4	12.1	10.9	12.1	11.0	12.1	10.9
15-19	9.6	9.7	9.7	9.2	9.6	9.1	9.6	9.2
20-24	7.2	8.2	7.4	8.3	7.5	8.3	7.5	8.2
25-29	6.4	7.5	6.2	7.4	8.2	7.4	6.2	7.4
30-34	5.4	6.4	5.8	6.7	5.8	6.7	5.7	6.7
35-39	5.5	6.2	5.2	5.8	5.2	5.9	5.2	5.9
40-44	4.6	4.9	4.7	5.0	4.7	4.9	4.7	5.0
45-49	4.2	4.1	4.1	4.1	4.1	4.1	4.1	4.1
50-54	3.4	3.3	3.3	3.2	3.3	3.2	3.3	3.2
55-59	2.6	2.4	2.6	2.5	2.6	2.5	2.6	2.5
60-64	2.2	2.1	2.0	1.8	2.0	1.8	2.0	1.8
65-69	1.2	1.1	1.4	1.3	1.4	1.3	1.4	1.3
70-74	1.0	1.1	1.0	1.1	1.0	1.1	0.9	0.9
75-79	0.6	0.5	0.6	0.5	0.6	0.5	0.6	0.7
80+	0.9	0.9	0.9	1.0	0.9	1.0	0.9	1.0
Index of dissimilarity with reported	-	-	0.95	1.50	0.95	1.40	1.55	2.50

Table 6: Reported and smoothed age-sex distributions and indices of dissimilarity for Kenya, 1969

Age Group	Reported		CF Smooth		QD Smooth		OS Smooth	
	M	F	M	F	M	F	M	F
0-4	19.3	19.2	19.3	19.3	19.4	19.3	19.8	20.1
5-9	16.8	16.4	16.7	16.4	16.7	16.4	16.2	15.4
10-14	13.0	12.2	13.0	12.2	13.0	12.3	13.0	12.2
15-19	10.2	10.0	10.3	9.9	10.3	9.8	10.2	10.0
20-24	7.8	8.2	7.9	8.6	7.9	8.6	7.9	8.6
25-29	6.4	7.5	6.3	7.1	6.2	7.1	6.3	7.2
30-34	5.1	5.5	5.3	5.7	5.3	5.7	5.3	5.7
35-39	4.6	4.8	4.4	4.6	4.4	4.6	4.4	4.6
40-44	3.5	3.7	3.7	3.7	3.7	3.7	3.7	3.7
45-49	3.1	3.0	3.0	3.0	3.0	3.0	3.0	3.0
50-54	2.4	2.5	2.5	2.4	2.5	2.4	2.5	2.4
55-59	2.1	1.9	2.0	2.0	2.0	2.0	2.1	2.0
60-64	1.9	1.7	1.8	1.7	1.8	1.6	1.8	1.6
65-69	1.4	1.2	1.4	1.2	1.4	1.3	1.5	1.3
70-74	0.9	0.8	0.9	0.8	0.9	0.8	1.0	0.9
75-79	0.6	0.5	0.6	0.5	0.6	0.5	0.4	0.4
80+	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9
Index of dissimilarity with reported	-	-	0.78	0.75	0.75	0.95	1.30	1.85

Table 7: Reported and smoothed age-sex distributions and indices of dissimilarity for Lesotho, 1976

Age group	Reported		CF Smooth		QD Smooth		OS Smooth	
0-4	14.7	13.7	14.7	13.6	14.8	13.5	14.0	12.9
5-9	13.4	12.4	13.4	12.4	13.4	12.4	13.9	13.2
10-14	13.2	12.5	12.6	12.4	12.5	12.3	12.6	12.4
15-19	10.2	10.7	10.8	10.9	10.9	11.0	10.8	10.8
20-24	8.4	8.9	8.5	8.6	8.5	8.7	8.6	8.8
25-29	7.1	6.8	7.0	7.0	7.0	7.0	6.9	6.9
30-34	5.6	5.5	5.8	5.5	5.8	5.6	5.8	5.5
35-39	5.2	4.8	5.1	4.8	5.0	4.7	5.1	4.8
40-44	5.3	5.2	4.9	4.7	4.9	4.7	4.8	4.7
45-49	4.0	3.8	4.4	4.3	4.4	4.3	4.4	4.3
50-54	3.3	3.3	3.8	3.7	3.7	3.7	3.8	3.7
55-59	3.5	3.5	3.0	3.1	3.0	3.1	3.0	3.1
60-64	2.1	2.5	2.1	2.4	2.1	2.4	2.2	2.4
65-69	1.5	1.8	1.5	2.0	1.5	2.0	1.5	1.9
70-74	1.0	1.6	1.0	1.6	1.0	1.6	1.0	1.8
75-79	0.9	1.6	0.9	1.6	0.9	1.6	1.0	1.6
80+	0.6	1.4	0.6	1.4	0.6	1.4	0.6	1.4
Index of dissimilarity with reported	-	-	1.80	1.45	1.90	1.60	2.55	2.05

Table 8: Reported, adjusted and projected female proportional age distribution for Swaziland

Age group	1976 Reported (%)	1976 Adjusted (%)	1981 Projected (%)	1986 Projected (%)
0 - 4	17.0	17.6	19.0	19.0
5 - 9	15.1	14.7	14.1	15.3
10 -14	12.8	12.6	12.3	11.8

Source: Kingdom of Swaziland (Central Statistical Office), Report on the 1976 Swaziland Population Census, Vol. I and III, 1979-80.

Table 9: Comparison of adjusted and projected population distributions by using logit method and UN adjustment techniques to female population, Botswana 1981.

Age Group	Reported (%)	Logit		UN Method	
		Adjusted (%)	1986 Projected (%)	1991 Projected (%)	Adjusted (%)	1986 Projected (%)	1991 Projected (%)
0-4	17.3	18.4	17.8	16.9	18.1	18.6	18.1
5-9	15.1	14.0	15.0	15.0	15.0	15.0	15.4
10-14	12.3	11.7	11.9	13.1	12.3	12.6	12.6

Source: Republic of Botswana (Central Statistical Office), Draft report on the Analysis of the 1981 Census (ECA communication) and Report on the Analysis of 1981 Census, Gaborone, 1985.

Table 10: Reported and adjusted age-sex distribution for Egypt

Age Group	1976 Reported % Male	Female	1976 Adjusted % Male	Female	Sex ratios 1976	1980	1950
0-4	13.6	13.9	14.6	14.5	97.8	104.4	102.6
5-9	13.0	12.6	13.0	12.7	103.2	106.2	100.9
10-14	13.8	12.9	11.7	11.4	107.0	106.4	101.0
15-19	11.4	10.3	10.3	10.2	110.7	104.7	101.5
20-24	8.2	8.7	9.0	9.0	94.3	103.7	102.2
25-29	7.1	7.6	7.8	7.9	93.4	102.4	106.9
30-34	5.6	6.1	6.8	6.9	91.8	102.2	109.1
35-39	5.5	5.7	5.8	6.0	96.5	100.2	111.8
40-44	5.0	5.3	4.9	5.1	94.3	99.6	123.9
45-49	4.2	4.1	4.1	4.3	102.4	98.9	499.2
50-54	3.9	4.1	3.4	3.6	95.1	97.9	499.2
55-59	2.6	2.3	2.8	2.9	113.0	100.1	593.2
60-64	2.6	2.7	2.1	2.2	96.3	99.0	680.5
65-69	1.5	1.3	1.6	1.6	115.4	97.6	781.6
70-74	1.1	1.3	1.0	1.0	84.6	103.7	975.0
75-79	0.5	0.7	0.7	0.5	71.4	145.2	1159.0
80+	0.4	0.4	0.4	0.1	100.0	414.8	1726.8
Total	100.0	100.0	100.0	100.0	103.7	103.7	124.9

Table 11: Evaluation of adjusted data by using PES information - Liberia, 1974 (males)

Age group	0-4	5-9	10-14	15-19	20-24	25-29	30-34	35-39	40-44
C(a)	15.18	30.28	41.81	51.77	56.94	66.02	72.22	78.46	83.22
C(a)§	15.39	30.55	42.14	52.75	60.13	67.21	73.45	79.65	84.21
CBR	39.77	43.93	43.30	42.71	40.30	39.35	38.45	38.81	38.14
CBR§	40.39	44.43	43.80	44.07	41.83	40.90	40.13	40.63	39.98
CBR (Mid Range)				4.53			3.41§		
Sum of squared deviations				.098			.081§		

§ These pertain to adjusted values.

Source: Ewbank, D.C., Age misreporting and age selective underenumeration;

Sources Patterns and Consequences for Demographic Analysis, National Academy Press, Washington, D.C. 1981.

Outline of the Ratio (CF) Quadratic (QF), Osculatory (OS) and 5 point moving average (UN) smoothing procedures

To smooth a population distribution of 5-year age groups, the CF, QD and OS procedures first group the 5-year age groups into 10-year age groups before separating the 10-year age groups into 5-year age groups.

(i) The Carrier-Farrag (CF) ratio method assumes that there exists a relationship between the population in 5-year age groups such that

$$5^{P}x = K_{(5^{P}x + 5)} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (1)$$

where $5^{P}x$ = population in the age group x, x+9

Then the population in a 10-year age group is given by

$$10^{P}x = 5^{P}x + 5^{P}x+5 = K_{(5^{P}x+5)} + 5^{P}x+5$$

$$= 5^{P}x+5\ (1 + K) \quad \ldots\ldots\ldots\ldots \quad (2)$$

Hence

$$5^{P}_{x+5} = 10^{P}x/(1 + K) \quad \text{and} \quad 5^{P}x = 10^{P}x - 5^{P}x+5$$

Assuming that the four 5-year age groups between ages X-10 and X+10 have the same average rate K, the estimation of K is made as follows:

$$K = \left(\frac{10^{P}x-10}{10^{P}x+10} \right) 1/4 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (3)$$

(ii) According to the quadratic procedure,

$$5^{P}x = 1/2 \left(10^{P}x\right) + \frac{1}{16} \left(10^{P}x-10 - 10^{P}x+10\right) \quad \ldots\ldots\ldots\ldots \quad (4)$$

and

$$5^{P}x+5 = 10^{P}x - 5^{P}x \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (5)$$

(iii) The Sprague Osculatory procedure uses two polynomials. A third degree polynomial is used for breaking down the first two and two last 10-year age groups. For intermediate 10-year age groups, a fourth degree polynomial is used. For the first age group, the formula is

$$5^{P}0 = E_1(10^{P}0) + E_2\ (10^{P}10) + E_3\ (10^{P}20) + E_4\ (10^{P}30) \quad \ldots\ldots \quad (6)$$

and for a central age group,

$$5P_x = C_1(10P_{x-20}) + C_2(10P_{x-10}) + C_3(10P_x) + C_4(10P_{x+10}) + C_5(10P_{x+20}) \quad \ldots\ldots\ldots\ldots(7)$$

where C_i and E_i = the coefficients 1/ for central and extreme age groups.

(iv) Regarding the UN procedure, the formula is

$$5P^S_x = 1/16 \; (-5P_{x-10} + 4\,5P_{x-5} + 10\,5P_x + 4\,5P_{x+5} - 5P_{x+10}) \quad \ldots\ldots\ldots\ldots\ldots\ldots(8)$$

where $5P^S_x$ = the smoothed age group x, x+4

and

$5P_x$ = the reported age group x, x+4.

1/ Details of these coefficients are contained in Arriage, E. et al, Computer Programmes for Demographic Analysis (Bureau of the Census, 1976) pp. 488-489.